

# Puce de génotypage Infinium<sup>®</sup> multi-espèces vs mono-espèce : bilan

Aurélie Bérard  
EPGV

**E P G V**

**Étude du Polymorphisme des Génomes Végétaux**

**US1279 - BAP**

# Pourquoi une puce multi-espèces: contexte

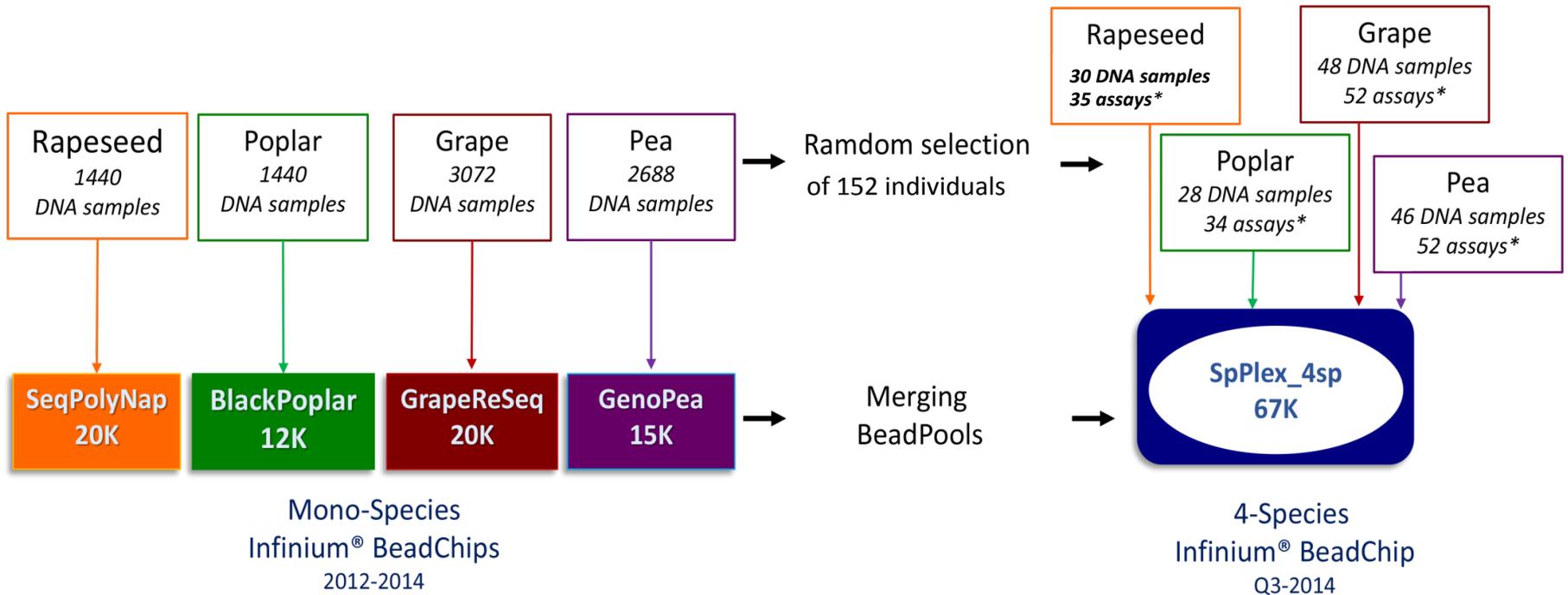
- Depuis 2005 le génotypage illumina Infinium® par puce d'hybridation donne des résultats très fiables pour des milliers de SNP sur des milliers d'individus d'organismes très différents, incluant les plantes.
- Les données produites sont utilisées dans de nombreuses études et donnent de nouvelles perspectives notamment pour la génétique de population et les analyses de diversité
- Cet outil pourrait être utilisé dans de nombreuses applications: le contrôle d'individu, les approches de génétique assistée par marqueur...
- Cependant ces puces restent un outil assez peu flexible et coûteux surtout si on a peu d'individus: le coût initial et le nombre minimal d'individus reste le facteur limitant à l'utilisation des ces puces.

# Pourquoi une puce multi-espèces: contexte

- Il existe déjà une alternative pour réduire les couts: les puces « add-on »: puces catalogue animales et récemment végétale (Maïs) sur lesquelles on peut rajouter un set de SNP mais il faut toujours un nombre minimal d'individus pour la première commande (1152).
- Une puce multi-espèces pourrait être une autre alternative pour par exemple:
  - Permettre d'accéder au génotypage quand on a un nombre limité d'individus en profitant du regroupement de plusieurs espèces
  - Diminuer les coûts en bénéficiant de la commande d'un grand nombre d'échantillons
- Nous avons donc validé le génotypage et évalué la fiabilité des résultats d'une puce de génotypage plante constituée de 4 espèces

# Description de la puce multi-espèces

- 4 espèces: colza, peuplier, vigne, pois



\*Assays : total number of genotyped samples (singletons, duplicates, triplicates)

# Design de la puce multi-espèces

From NGS resequencing to Infinium® genotyping : 1-species BeadChips



4-species BeadChips  
= SpPlex\_4sp

Project	NGS Re-seq	Infinium® Beadchip	Infinium® Beadchip	Infinium® Design	Infinium® Deliver	Infinium® Success	Infinium® Deliver
Coordinator	Sample #	Application	Name_Kbeads Sample #	SNPs #	SNPs # %	SNPs # %	SNPs # % vs 1-species
Grants							
<b>SeqPolyNap</b> B.Chaloub, INRA ANR-09-GEMN-021	5	Genome A/C scaffolds anchoring	SeqPolyNap_20K 1440 samples	17 607	15 932 90%	13 267 83%	15 363 96.4%
<b>BlackPoplar</b> P.Faivre Rampant, INRA EU-FP6-16322 : EVOLTREE project EU-FP7-211868 : NovelTree project	52	QTL & Diversity analysis	BlackPoplar_12K 1440 samples	10 331	9 127 88%	8 259 90%	8 736 95.7%
<b>GrapeReSeq</b> A.F.Adam Blondon, INRA ANR-08-KBBE-008	41	Diversity analysis	Grape_20K 3072 samples	20 000	18 071 90%	12 815 71%*	17 510 96.9%
<b>GenoPea</b> J.Burstin, INRA ANR-09-GMNE-026	16	Genetic map & Diversity analysis	Genopea_15K 2688 samples	15 000	13 204 88%	13 156 99.6%**	12 738 96.5%

*\*with « global » clusterfile for all samples, need to refine related to samples/populations*

*\*\*33% to 67% depending on RIL population*

- puce multi-espèces construite à partir de la même synthèse des 4 beadpools précédemment designés.
- A cause de la présence de mêmes illumina codes dans les 4 beadpools déjà existants la puce multi-espèces ciblant 54 347 SNP soit 96,5% de la totalité des marqueurs des 4 puces mono-espèce.

# Echantillons, Génotypage et SNP Allele calling

- ADN:

Pour valider le génotypage sur la puce multi-espèces nous avons utilisé:

- Un échantillon des mêmes ADN (même tube stock) génotypés sur chaque puce mono-espèce (sauf pour le colza: nouvel envoi des ADN pré-dilués à partir des tubes stocks restitués au partenaire)
- A défaut un autre tube stock d'une autre extraction du même individu (ADN vigne)
- Certains ADN ont été dupliqués, tripliqués...

- Génotypage et Allele calling:

Le génotypage a été réalisé sur BeadChips au format 24-samples

Le calling a été réalisé en appliquant directement pour chaque espèce le ou les clusterfile(s) préalablement établi pour chaque puce mono-espèce

Seuls les marqueurs ayant donné des résultats exploitables pour les puces mono-espèce ont été retenus pour l'analyse comparative

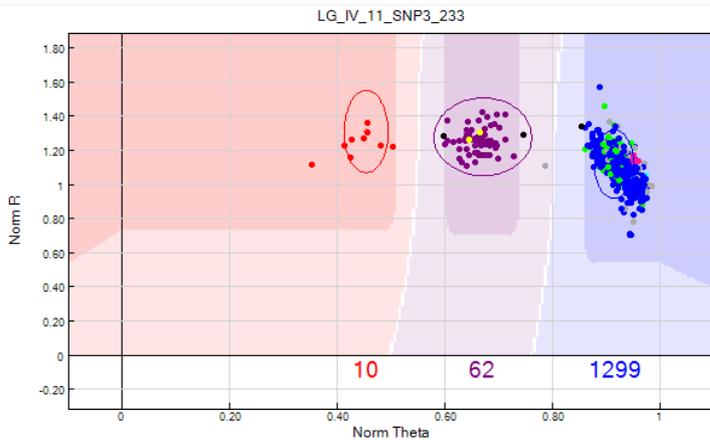


# Résultats de la puce multi-espèces: sans réajustement des clusterfiles

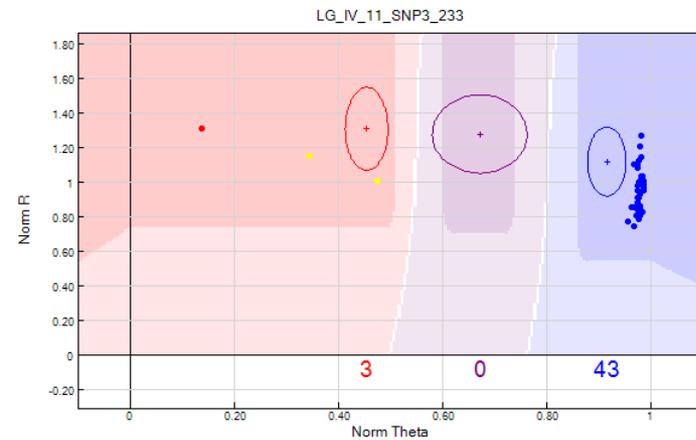
- Reproductibilité inter-puce moyenne > 99,8% sans les données manquantes
- Reproductibilité intra-puce des répétitions >99,99%
- Individus avec une faible reproductibilité:
  - Soit l'ADN est de mauvaise qualité et il a beaucoup de données manquantes
  - Soit l'ADN génotypé dans la puce multi espèce est différent de celui génotypé dans la puce mono-espèce (erreur d'étiquetage )
  - Clusterfile à ré-affiner du au nombre et au panel d'échantillons

# Analyse des discordances: quelques exemples chez le peuplier

## mono-espèce : génotype AB

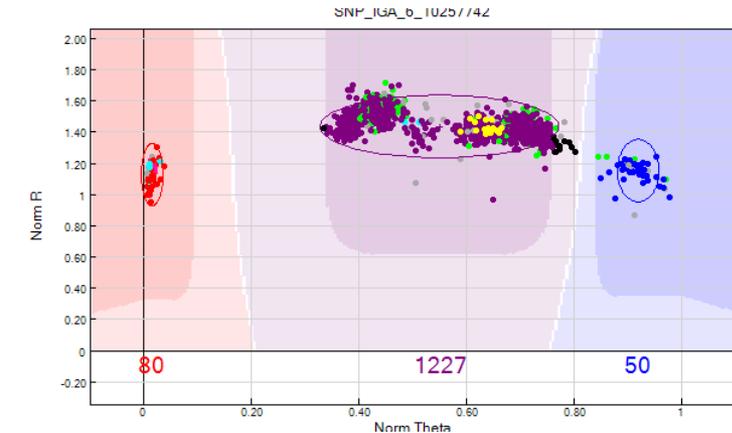


## multi espèce : génotype AA

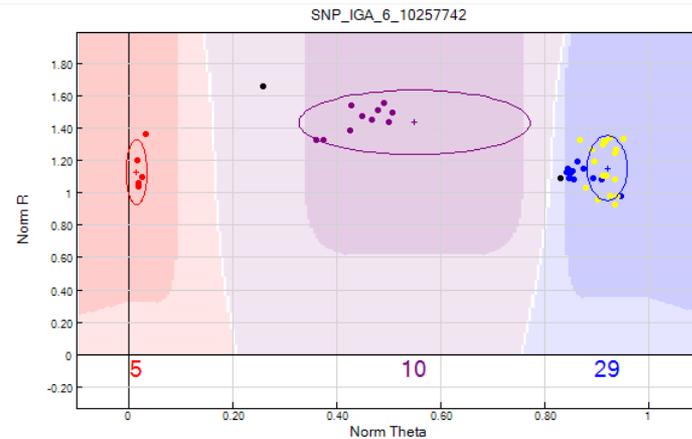


→ Réajustement du cluster file dans la multi-espèces

## mono-espèce : génotype AB



## Multi espèce : génotype BB

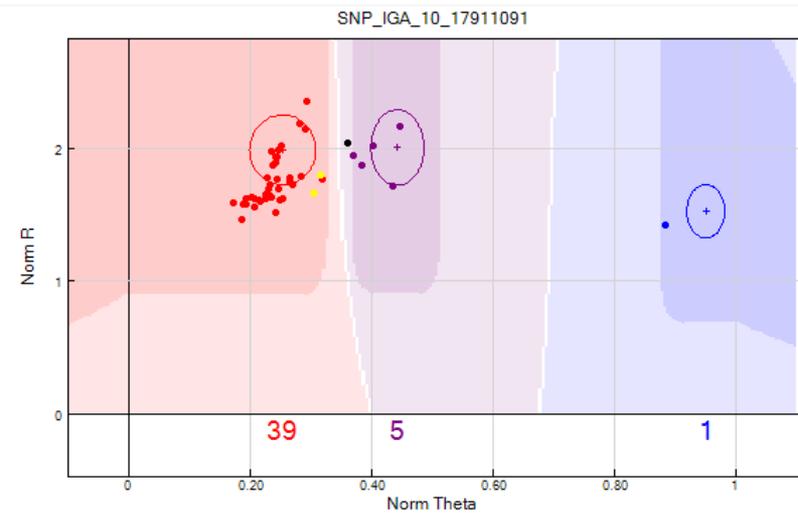
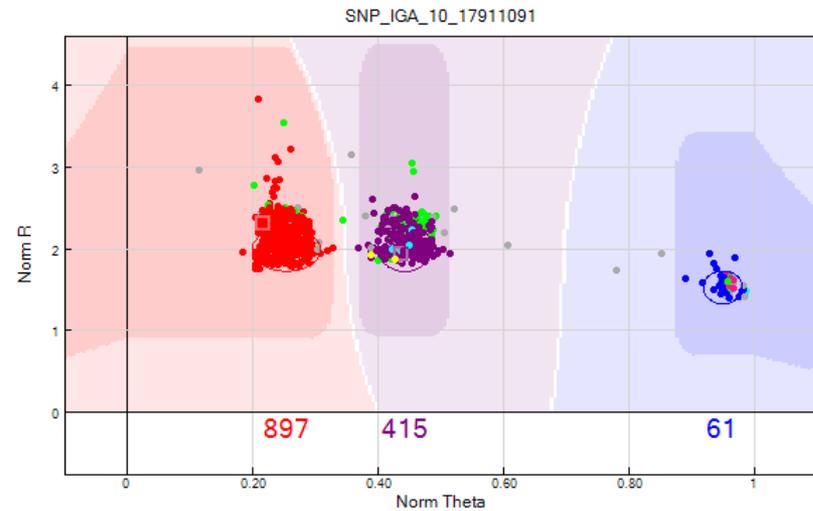
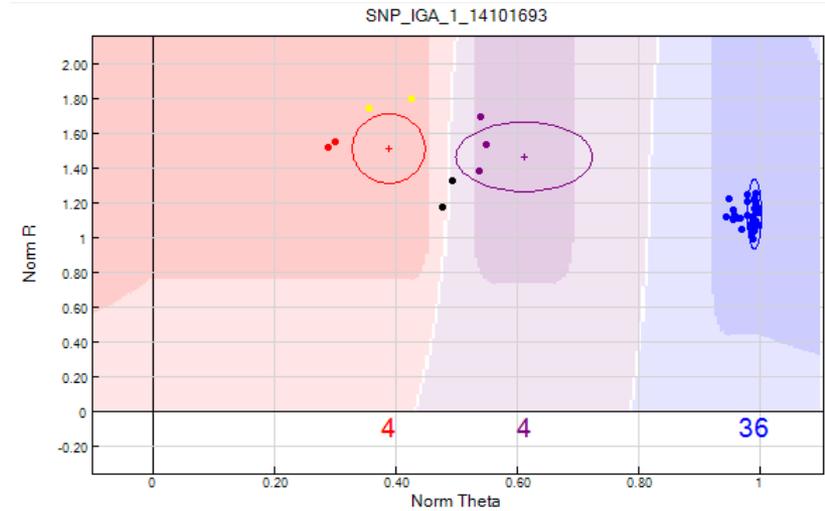
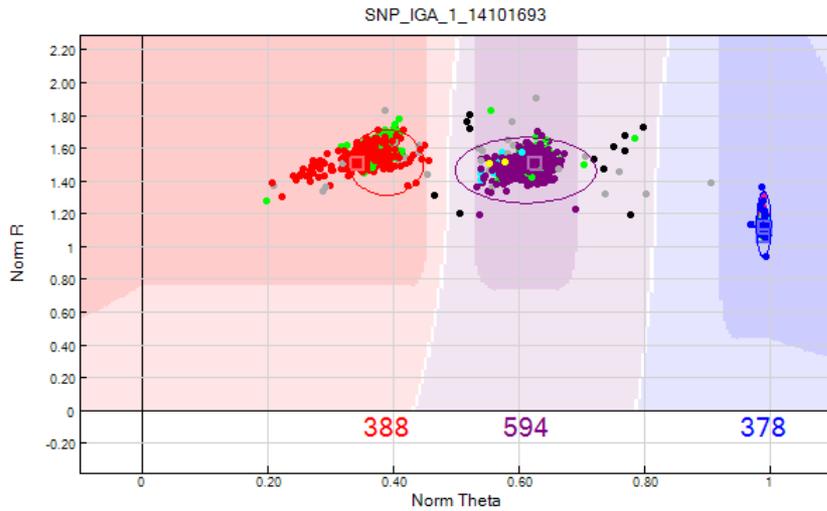


→ Problème de cluster file et de population dans la mono-espèce

# Analyse des discordances: quelques exemples chez le peuplier

mono-espèce : génotype AB

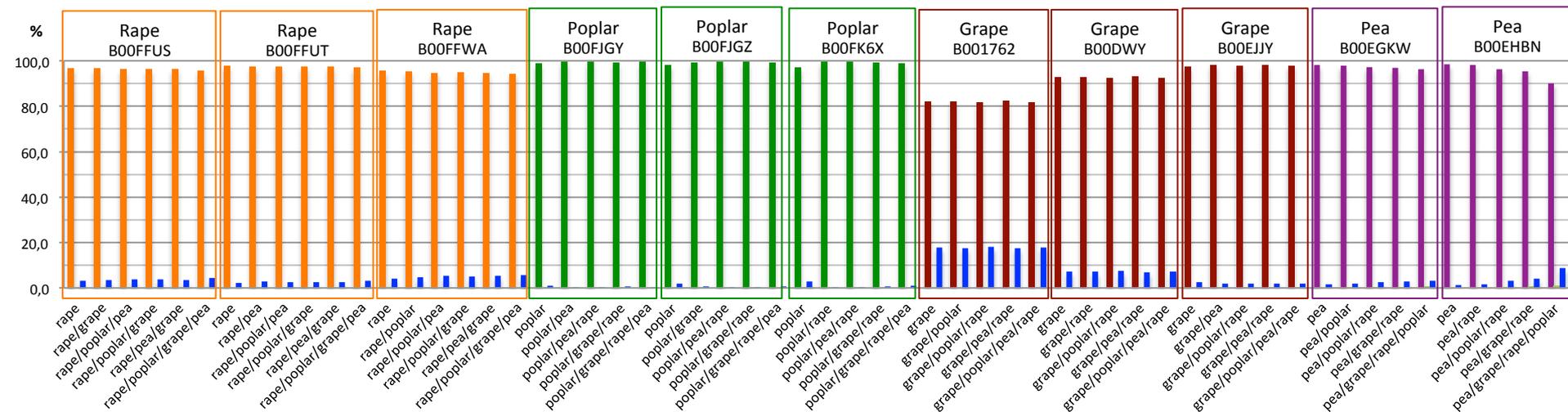
multi espèce : génotype AA



→ Vraies discordances

# Pour le Fun: analyse de mélanges d'ADN

- mélanges d'ADN de 2 à 4 espèces génotypés avec la puce multi-espèces



- Plus on mélange d'espèces plus la reproductibilité diminue mais reste variable selon le type de mélange
- Reproductibilité entre 99,4% et >99,9% selon les mélanges (sans les données manquantes)

# Conclusions et perspectives de la puce multi-espèces

## Avantages de la puce multi-espèces:

- Très haute reproductibilité des résultats entre la puce multi-espèces et la puce mono-espèce: > 99,8%
- Diminution des coûts par échantillon
- Partage de la contrainte du minimum d'individus par espèce ( minimum de première commande 1152 individus)
- Possibilité d'étaler le génotypage sur plusieurs années

## Contraintes:

- Le coût de la multi-espèces doit être inférieur à la mono-espèce
- Dans le cas de mélange d'ADN contrainte logistique que les échantillons de chaque espèce arrivent en même temps

## Exemples de perspectives d'utilisation d'une puce multi-espèces en groupant des projets/espèces:

- Contrôle d'individu:  $\leq 1\ 000$  SNP : uniquement possible en infinium® à partir de 3K SNP
- Nombre d'échantillons inférieur à 1152
- Puce générique à 10k SNP par projets/espèces pour environ 5 à 10 espèces qui nécessite de s'engager sur un nombre minimal de 10-15K échantillons sur 2 ans pour pouvoir négocier

# Remerciements

- Equipe EPGV  
Aurélie Chauveau  
Patricia Faivre-Rampant  
Marie Christine Le Paslier  
Elodie Marquand  
Dominique Brunel  
Aurélie Canaguier



- Illumina®  
André Eggen  
Yann Legros



- CNG  
Anne Bauland-Augé  
Céline Besse



- Les partenaires des 4 puces mono-espèce